A statistical model for regional tornado climate studies

Thomas. H. Jagger,* James B. Elsner, and Holly M. Widen

Department of Geography, Florida State University, Tallahassee, Florida, USA.

**Corresponding author address:* Department of Geography, Florida State University, 113 Collegiate Loop, Tallahassee, FL 32306, USA.

E-mail: tjagger@fsu.edu

ABSTRACT

Tornado reports are locally rare, often clustered, and of variable quality making it difficult to use them directly to describe regional tornado climatology. Here a statistical model is demonstrated that overcomes some of these difficulties and produces a smoothed regional-scale climatology of tornado occurrences. The model is fit to data aggregated at the level of state counties. These data are annual population, annual tornado counts and an index of terrain roughness. The model has a term to capture the smoothed frequency relative to the state average. The model is used to examine whether terrain roughness is related to tornado frequency and whether there are differences in tornado activity by County Warning Area (CWA). A key finding is that tornado reports increase by 13% for a two-fold increase in population across Kansas after accounting for improvements in rating procedures. Independent of this relationship tornadoes have been increasing at an annual rate of 1.9%. Another finding is the pattern of correlated residuals showing more Kansas tornadoes in a corridor of counties running roughly north to south across the west central part of the state consistent with the dryline climatology. The model is improved by adding terrain roughness. The effect amounts to an 18% reduction in the number of tornadoes for every ten meter increase in elevation standard deviation. The model indicates that tornadoes are 51% more likely to occur in counties served by the CWAs of DDC and GID as elsewhere in the state. Flexibility of the model is illustrated by fitting it to data from Illinois, Mississippi, South Dakota, and Ohio.

1. Introduction

Broadscale tornado climatology in the United States is well described and physically understood. The seasonal spread of the tornado threat from the deep South northward into the northern Plains and Midwest during summer is tied to the poleward migration of the jetstream (Brooks and Doswell 2001). A concentration of tornado activity across Oklahoma and Kansas during spring is linked to the vertical intersection of mid-level dry air from the Rockies and abundant low-level moist air from the Gulf of Mexico (Schultz et al. 2014).

Regional-scale tornado climatology is less well described and poorly understood. One reason is because tornadoes are discrete events, spatially clustered, and locally quite rare. Another reason is because of the variable quality of the available records (Diffenbaugh et al. 2008; Brooks 2013). While the U.S. tornado database is the largest in the world, it contains issues that limit its utility for climate studies (Doswell et al. 1999). For instance, improved observation practices have led to an increase in the reporting of weak tornadoes (Verbout et al. 2006; Doswell 2007). Even today many weak, short duration tornadoes likely go undocumented in places with few people or poor communication infrastructure. This observational effect is well known (Snider 1977; Doswell et al. 1999) although it appears to have diminished during the most recent decade (Elsner et al. 2013).

Various methods for quantifying and modeling the observational effects have been proposed (King 1997; Ray et al. 2003; Anderson et al. 2007). Most studies assume a uniform region of activity and estimate tornado frequency within a subset of the region likely to be most accurate. The uniform regions are defined by the available data. Tornado reports are often aggregated using kernel smoothing (Brooks et al. 2003; Dixon et al. 2011; Shafer and Doswell 2011, e.g.,). Spatial density maps that show regions of higher and lower tornado frequency are useful for exploratory analysis and hypothesis generation but are less so for modeling since the choice of kernel bandwidth is subjective. Another drawback is the implicit assumption that tornado occurrences are independent. This is generally not the case as a single supercell thunderstorm can generate a family of tornadoes (Doswell and Burgess 1988).

This research asks the question; how can regional tornado climatology be recovered from a heterogeneous database of rare, clustered events? The question is answered with a statistical model that produces a map of smoothed tornado occurrence reflecting regional patterns of possible physical forcing. The available data are first aggregated to the county level. Aggregation makes it easy to leverage human and environmental data (population, terrain, percent agriculture, etc.) in attempts to control for known effects in the data. The model is fit using the method of integrated nested Laplacian approximation (INLA) to solve the Bayesian integrals. This setup accommodates non-normally distributed counts and a correlated random-effects term. The random-effects term shows where tornado activity is high relative to the state average. The method described in this paper is valuable because it has the potential to uncover the 'true' spatial pattern of tornado activity and it provides a solid foundation for statistical tests about the relationship between tornadoes and climate.

The data preparation and modeling procedures are described first for Kansas. The procedures are then demonstrated for Illinois, Mississippi, South Dakota, and Ohio representing different tornado-prone areas in the United States. For each state an index of terrain roughness is tested to see whether it improves the model fit. Similarly the National Weather Service (NWS) County Warning Areas (CWA) are used to identify areas with significantly higher and lower tornado rates.

The balance of the paper is outlined as follows. The tornado database and identified reporting issues are described in section 2. The tornado report frequency by Kansas county is evaluated in section 3. The statistical model used to estimate tornado occurrence by county while controlling for non-physical factors is described in section 4 and the results from fitting the model to tornado reports first from Kansas then from Illinois, Mississippi, South Dakota, and Ohio are shown in section 5. The influence of terrain roughness on tornado frequency conditional on the model is examined in section 6. In section 7, key findings are summarized and suggestions made for future work. The code to produce the table and all the figures is available at `http://myweb.fsu.edu/jelsner/StateTornadoModel.html`.

2. Data Preparation

a. Boundaries, elevation, and population

The model is written with the open-source R language using freely-available government data including tornadoes from the U.S. Storm Prediction Center (SPC), population and administrative boundaries from the U.S. Census Bureau, and elevations from NASA's Shuttle Radar Topography Mission (SRTM). The data are prepared as follows. First county administrative boundaries for the United States are downloaded and read into R as vector polygons from `https://www.census.gov/geo/maps-data/data/cbf/cbf_counties.html` at a resolution of 1:5 million and subset by the state of interest using the Federal Information Processing Standard (FIPS) code. Then digital elevation model (DEM) data are downloaded from `http://www.viewfinderpanoramas.org` at a resolution of three arc seconds (approximately 80 m) and read into R as a raster. The elevation raster is cropped to the state boundary. Next CWA labels from `http://www.nws.noaa.gov/geodata/catalog/wsom/data/bp03de14.dbx` are attached to each county. The results for Kansas are displayed on a map in Fig. 1.

Elevation (above mean sea level) ranges from less than 220 m in the east to higher than 1220 m in the west. The Kansas River in the northeast and its tributaries extending westward are visible at this spatial resolution. These elevations are used to compute an index of terrain roughness. The three-letter abbreviation of the corresponding CWA is given in each county. The CWAs include Dodge City (DDC), Goodland (GLD), Topeka (TOP), Wichita (ICT), North Platte (LBF), Omaha/Valley (OAX), and Kansas City/Pleasant Hill (EAX). The DDC NWS is responsible for 27 Kansas counties followed by 26 for ICT and 23 for TOP.

Data preparation continues by adding annual population estimates over the period 1970–2012 from `http://www.nber.org/data/census-intercensal-county-population.html` to each county. The percentage change over this period using 2012 as the baseline is displayed on a Lambert conformal conic map in Fig. 2. Counties in blue indicate more people in 2012 compared to 1970. Counties to the south and west of Kansas city show the largest increases. Butler and Sedgwick counties (Wichita area) and Ford, Gray, and Finney (Dodge City area) also show large percentage increases although the latter area has fewer people (Fig. 3). Population densities exceeding 190 people per square kilometer are found in Wyandotte (Kansas City), Johnson, and Sedgwick counties.

b. Tornado Tracks

Next the SPC database containing all reported tornadoes in the United States over the period 1950–2013 is obtained from `www.spc.noaa.gov/gis/svrgis/zipped/tornado.zip`. Individual reports in the database are compiled by the NWS offices and reviewed by the National Climate Data Center (Verbout et al. 2006). The database comes in a shapefile format with each tornado provided as a straight line track. Tornado information in the database is considered reliable for climate studies (Ramsdell and Rishel 2007). The tornado track is the great circle line (no width) between the estimated start (touchdown) and end locations. Locations are recorded with two digit decimal precision prior to 2009 and four digit afterwards. Locations are more accurate later in the record when estimates are made with GPS. Not all tornadoes track in a straight line nor do they all remain in contact with the ground along the entire path. No attempt is made to adjust for possible variations from a continuous straight line track.

Tornado reports tend to be more numerous near cities compared to rural areas but this spatial variation is decreasing with time (Elsner et al. 2013). Moreover, improvements in observational practices tend to result in a larger number of tornado reports, especially reports of weak tornadoes (Doswell et al. 2005; Verbout et al. 2006). Tornadoes are rated on a damage scale from 0 (least) to 5 (Fujita and Pearson 1973; Edwards et al. 2013), with the earliest tornadoes in the database rated retroactively (Schaefer and Edwards 1999; Anderson et al. 2007; Coleman and Dixon 2014). To improve the precision on the ratings the Enhanced Fujita Scale, which includes more damage indicators, was adopted in 2007 (Potter 2007). Changes to population and to the rating procedures result in a heterogeneous database. Consistent with advice given by the SPC (Verbout et al. 2006) our analysis is limited to tornadoes rated EF1 and higher on the damage scale. In this paper the word 'tornado' refers to tornadoes that received a damage rating of at least EF1. County tornado counts are accumulated for each tornado track that falls within or that crosses into the county for each year.

The result is a space-time database with constant-time attributes that include county area and terrain roughness and variable-time attributes that include the annual number of tornadoes and population density. Area is converted to units of square kilometers and the tornado rate per county is computed as the number of tornadoes per 10,000 square kilometers per year. Tornadoes are most numerous across central Kansas (Fig. 4). The larger counties tend to have more tornado reports although the relationship is not large [r = .34 (.19, .48) 90% CI] since the counties tend to have similar sizes. Regional hotspots include Sedgwick County (city of Wichita) and parts of the northeast in the counties around Kansas City. The correlation between the 2012 county population and the number of tornadoes is positive but weak [r = .04 (−.12, .20) 90% CI]. The annual number of tornado reports for the state as a whole has increased since 1970 at a rate of less than one per year, but the trend is not significant (Fig. 5). Summary statistics are listed in Table 1.

3. Model for County-Level Tornado Counts

The main idea of this paper is a model for tornado occurrence at the county level. The model is more useful for climate studies than are the raw counts because it includes a term that captures the smoothed frequency relative to the state average after accounting for known non-climate factors. To account for changes in tornado reporting due to population shifts over time the $\log_2$ annual county population density is included as a fixed-effect term. Further, to account for improvements in rating procedures over time, the calendar year and an interaction term of year with $\log_2$ population density are also included as fixed-effect terms. Finally to account year to year changes a random effect term was added.

Inferences on the number of tornadoes in each county, s for each year t, $T_{s,t}$ is assumed to be adequately described by a negative binomial distribution (Elsner and Widen 2014) with parameters probability p and size n. If X is a random sample from this distribution, then the probability that $X = k$ is $P(k|r,p) = \binom{k+r-1}{k}(1-p)^r p^k$, for $k \in 0,...,\infty$, $p \in (0,1)$ and $r > 0$. This relates the probability of observing k successes before the r failure of a series of independent events with probability of success equal to p.

The distribution is generalized by allowing r to be any positive real number and it arises from a Poisson distribution whose rate parameter has a gamma distribution. Whereas the Poisson distribution has a variance equal to its mean, the negative binomial distribution is over dispersed. That means the ratio of the variance to the mean exceeds one implying that the underlying process that generates the counts is clustered. To simplify inferences, the distribution is re-formulated using the mean, $\mu = r\frac{p}{1-p}$ and the size r which allows a separation of the mean effect from the dispersion.

The mean of the negative binomial distribution, $\mu_{s,t}$ is linked to a structured additive response $\nu_{s,t}$ through the link-function and normalized area offset, A_s as $\log(\mu_{s,t}/A_s) = \nu_{s,t}$ The dispersion is managed using a normalized size parameter n where the county size parameter is $n = r_{s,t}/A_s$ giving a dispersion of $1/p_{s,t} = 1 + \mu_{s,t}/n = 1 + \exp(\nu_{s,t})/n$ that depends only on the tornado density and n. To make n manageable the area of each county in square km is divided by 2000.

More concisely the model is:

$$
\begin{aligned}
T_{s,t}|\mu_{s,t}, r_{s,t} &\sim \text{NegBin}(\mu_{s,t}, r_{s,t}) \\
\mu_{s,t} &= \exp(A_s \nu_{s,t}) \\
\nu_{s,t} &= \beta_0 + \beta_1\,\text{lpd}_{s,t} + \beta_2\,(t - t_0) + \beta_3\,\text{lpd}_{s,t}(t - t_0) + u_s + v_t \\
r_{s,t} &= nA_s
\end{aligned}
$$

where NegBin($\mu_{s,t}$, $r_{s,t}$) indicates that the conditional tornado counts ($T_{s,t}|\mu_{s,t}, r_{s,t}$) are described by a negative binomial distribution with mean $\mu_{s,t}$ and size $r_{s,t}$, $\text{lpd}_{s,t}$ represents the base 2 logarithm of the annual population density for each region, and t_0 is the base year set to 1991 (middle year of the record).

The correlated spatial random effects term u_s follows an intrinsic Besag formulation with the sum to zero constraint (Besag 1975):

$$
\begin{aligned}
u_i|u_j, j \neq i, \tau &\sim N\left(\frac{1}{m_i}\sum_{i\sim j} u_j, \frac{1}{m_i}\tau\right) \\
\sum_{\forall i} u_i &= 0
\end{aligned}
$$

where N is the normal distribution with mean $1/m_i \cdot \sum_{i\sim j} u_j$ and variance $1/m_i \cdot 1/\tau$ where m_i is the number of neighbors of county i and τ is the precision; $i \sim j$ indicates the two counties i and j

are neighbors. Neighboring counties are determined by contiguity (queen's rule) using functions from the **spdep** package (Bivand 2014). The annual uncorrelated random effect, v_t, is modeled as a sequence of normally distributed random variables, with mean 0 and variance $1/\tau'$

The prior on the vector of spatial random effects is statistically independent from the vector of annual random effects. For each posterior sample the vector of spatial random effects has the same values for all years and the vector of annual random effects has the same values for all regions as implied by the subscripts in the model notation. Gaussian priors with low precision are assigned to the β's. To complete the model the scaled size (n) is assigned a log-gamma prior and the precision parameters (τ and τ' are assigned a log-Gaussian prior. Although yet to be used on county-level tornado data, a similar model was recently constructed for modeling hurricane data (Elsner and Jagger 2013) and these types models are frequently used for mapping disease rates (Schrödle and Held 2011; Blangiardo et al. 2013).

The priors and the likelihood are combined with Bayes rule to obtain the posterior distributions for the model parameters. Since the integrals cannot be solved analytically, a common technique is to use a Markov chain Monte Carlo (MCMC) algorithm to obtain samples from the posterior distributions. Here the method of integrated nested Laplace approximation (INLA) is used instead. INLA provides a fast alternative for models with a latent Gaussian structure (Rue et al. 2009) and is accomplished with functions from the **INLA** package (Rue et al. 2014).

4. Results

a. Fit, adequacy, and fixed effects

The model above is fit to the county-level tornado counts. The Deviance Information Criterion (DIC) is used as relative measure of how good the model fits the data. Versions of the model with and without the correlated random-effects term are compared. The DIC for the model that includes the correlated random-effects term is 5990 which compares with a DIC of 6027 for the model without it. The smaller the DIC, the better the model fit. The correlated random-effects is important to the model and is kept.

The model fits the data well. The probability integral transform (PIT) values modified for small counts are adequately described by a uniform distribution (Czado et al. 2009). The adequacy is checked by noting that the p-value on an Anderson-Darling (AD) goodness-of-fit test under the null hypothesis of a uniform distribution exceeds .15. The predictive quality of the model is assessed by the cross-validated log score. A smaller value of the score indicates better predictive quality (Gneiting and Raftery 2007). The log score is .635 for Kansas, which is better than the log scores for seasonal tornado models (Elsner and Widen 2014). The Brier score is .570 as the mean squared difference between the predicted probability and the actual count in each county for each year ($105 \times 45 = 4725$ predictions). The Brier score for the null model is .603

The coefficient on the logarithm (base 2) of population density has a posterior mean of .1187 [(.0655, .1723) 90% credible interval (CI)] (Table 1). This translates to an 13% [(exp(.1187) - 1) $\times$ 100%] increase for a doubling of the population. The coefficient on the year (trend) term has a posterior mean of .0189 [(.0054, .0323) 90% CI]; statistically significant and upward at a rate of 1.9% per year. The interaction term is also statistically significant with a posterior mean of $-.0045$, indicating a decrease in the influence of population density. In fact the model indicates that the influence of population density on the tornado reports will reach zero by the year 2017 [$\beta_1 + \beta_3\,(2017 - 1991) \approx 0$].

b. Correlated random effects

The random-effects term is the spatially correlated set of county-level residuals that provides a description of tornado occurrence statewide that accounts for population changes, differences in exposure, and trend. A map of this term reveals where tornadoes are more likely relative to the state average (Figure 6). Values are the posterior means and are expressed as a percent difference from the state average. Counties with significantly (at the 90% level) higher and lower rates are outlined in bold. Uncertainty on the magnitude of these values is measured by the posterior standard deviation (Fig. 7). Standard deviations tend to be lower (precision higher) in counties with more neighbors (away from the state borders).

The map features a north-south axis of above-average activity across the west central part of the state with lower activity to the west (as found in Brooks et al. (2003)) and generally lower activity to the east. The axis of above-average activity in the north is shifted somewhat farther to the east. The four counties of Hodgeman, Edwards, Pawnee, and Stafford in south central Kansas have tornado activity that exceeds the average by at least 40% as do Jewell and Republic counties in the north.

Nearly three quarters of Kansas tornadoes occur from April through June. Surface low pressure in eastern Colorado to the lee of the Rockies in response to westerly winds aloft produce veering southeasterly surface winds across the state. These winds transport moisture up slope (Fig. 1) with deep convection initiating in western Kansas along the dryline. The dryline forms in the High Plains during spring and separates moist air originating over the Gulf of Mexico from dry air originating over the southwestern United States and high plateau of Mexico (Schultz et al. 2007). Initial thunderstorm organization results in discrete supercells east of the dryline along a roughly north-south axis. The discrete cells tend to merge into a mesoscale convective system across eastern Kansas after sunset reducing the threat for tornadoes. Strong winds, heavy rains, and frequent lightning become the main concern to life and property.

c. Index of terrain roughness as a fixed effect

Next the model is used to test whether terrain roughness can help explain the pattern of tornadoes across Kansas. The test is motivated by the physical hypothesis that a tornado is somewhat more likely to occur, all else being equal, where the low-level inflow is unimpeded. Studies have shown that surface roughness affects this inflow; in particular it affects the velocity distribution, pressure distribution, and the core radius of the flow (Lewellen 1962; Davies-Jones 1973; Dessens 1972; Leslie 1977). An increase in terrain roughness causes the maximum tangential velocity to decrease (Leslie 1977). But experimental studies have argued that the roughness used in these studies are outside the range of values encountered in nature (Church et al. 1979).

Here the standard deviation in the 80-m resolution elevation data is computed within each county and used as a proxy for terrain roughness. Counties with smaller elevation standard deviations are smoother. Values range from a low of 11.3 m to 73.4 m with the smoother counties in the southeast part of the state. The model is refit using terrain roughness as an additional fixed effect. The DIC decreases to 5980 indicating a better model with this term included (Table 1). Elevation itself is not a significant term when included in the model.

The magnitude of the effect is indicated by the size of the coefficient. The posterior mean of the coefficient is $-.0186$ [($-.0268$, $-.0106$) 90% CI] indicating an 18% reduction in the tornado occurrence for every ten meter increase in elevation standard deviation. The significance of the effect is indicated with a plot of the posterior density (Fig. 8). The density is offset to the left of zero, where zero indicates the proxy for terrain roughness has no relationship to tornadoes at the county level.

This finding is consistent with Karpman et al. (2013) who show a negative relationship between the occurrence location of tornadoes and elevation variance. However, Karpman et al. (2013) consider only touchdown locations of intense (EF3+) tornadoes and a domain that covers the eastern two-thirds of the United States. They also use a coarser (approximately 1 km) elevation database.

Since the roughness term is significant it is added to the model and the correlated random-effects term re-evaluated (Fig. 9). The overall pattern remains unchanged with a corridor of enhanced activity across the west-central part of the state. This example shows how to test hypotheses concerning factors that could be related to tornado activity by representing the values at the county-level and included the term in the model.

d. County Warning Area as a fixed effect

Next the model is used to check whether there are significant variations in tornado activity by CWA. Variations do not necessarily imply different warning and verification practices. Nevertheless to improve consistency across offices it is instructive to know whether more attention to variations is warranted. The CWA term is treated as a factor variable where each county is given the name of the corresponding CWA (see Fig.1). The term is included as a fixed effect. The DIC with this term increases to 5981 indicating there is no significant pattern of tornado activity correlated to the arrangement of the seven CWAs over the state. However, when the DDC CWA (Dodge City, KS) and the GID CWA (Grand Island, NE) are included as a single combined binary variable (DDC and GID or neither) the DIC drops to 5977. The coefficient on the binary term is .4112 [(.2185, .6011) 90% CI] indicating that tornadoes are 51% more likely to have occurred in counties served by these two CWAs as elsewhere in the state. The DDC and GLD offices are responsible for warnings across central Kansas where tornadoes tend to be most numerous and the spatial random effect is mostly positive.

e. Illinois, Mississippi, South Dakota, and Ohio

The flexibility of the model is demonstrated by fitting it to data from four additional states including Illinois, Mississippi, South Dakota, and Ohio. The choice of states is based on a representative sample of other tornado-prone areas in the United States. The summary and model statistics discussed below are listed by state in Table 1. Maps of raw tornado counts by county for the four states are shown in Fig. 10. The procedures for preparing the data at the county level are the same as before. An exception occurs for South Dakota where an additional raster of elevations is needed for counties north of 45° N latitude. Like in Kansas, tornado counts are significantly correlated with county size in Illinois, Mississippi, and Ohio. South Dakota is the exception where the larger counties in the western half of the state tend to have fewer tornadoes compared to the smaller counties in the southeast corner.

Counties with more people also tend to have more tornado reports. This is particularly true for Mississippi which has a correlation between tornado frequency and population of .49 [(.34, .62) 90% CI] and for South Dakota which has a correlation of .39 [(.20, .55) 90% CI]. The pattern of tornadoes across Illinois features a diagonal axis of high frequency from southwest to northeast similar to the pattern noted in Wilson and Changnon (1971). However, this axis coincides with larger and more densely populated counties compared to the state average. The pattern of tornadoes in Mississippi features a hotspot in the vicinity of the city of Jackson. Across Ohio tornadoes are notably fewer in the mountainous regions of the southeast. Marginally significant downward trends in statewide tornado frequency are noted for South Dakota and Ohio (Fig. 11). A slight increase in the number of tornadoes is noted in Illinois and Mississippi since 2000.

Population density is a significant term in each of the models with South Dakota having the largest effect with a 28% increase in tornado reports for a doubling of the population. Mississippi is next with a 20% increase in tornado reports for a doubling of the population. Population is only marginal significant for Ohio. A significant downward trend at a rate of 1.7% per year is noted in the model for South Dakota tornadoes and a significant upward trend at a rate of 2.4% per year is noted in the model of Mississippi tornadoes. No significant upward trends are noted for tornadoes in Illinois and Ohio. The interaction term is significant for Mississippi and Ohio, marginally so for Illinois, and not significant for South Dakota.

Maps showing the correlated random effects from the state models are shown in Fig. 12. Illinois features a band of significantly below average frequency across the northern quarter of the state with much of the rest of the state above average. Some significant hotspots of above normal activity are noted across the midsection and over the extreme south. Mississippi shows a similar pattern with below normal frequency in the north and higher than average frequency across central and southern parts of the state. These north-south gradients are partially hidden in the map of raw counts but becomes conspicuous when controlling for county size and population density. The gradients are consistent with what would be expected over the long-term as the tornado season is longer in the south. South Dakota shows a well-defined mainly east-west gradient with significantly more tornadoes across the southeast and significantly fewer tornadoes in the west. Ohio features significantly fewer tornadoes across the southeast and a band of significantly more tornadoes running from near the city of Canton westward to the state line. The model with a correlated random-effects term is a type of smooth algorithm that accounts for population changes, differences in exposure, and trends.

Terrain roughness is a significant factor in the model for Mississippi tornadoes and marginally so for South Dakota but not elsewhere (Fig. 13). Like Kansas the significant coefficients are negative indicating more tornadoes with smoother terrain. The magnitude of the effect is a 10% reduction in Mississippi tornadoes for every ten meter increase in elevation standard deviation and a 2% reduction in South Dakota tornadoes for the same amount of increase in roughness. County-level elevation standard deviations range from 2 to 35 m in Mississippi and from 6 to 420 m in South Dakota. The CWAs are not a significant factor in explaining the pattern of tornadoes in Illinois and Ohio. However in Mississippi the JAN CWA (Jackson, MS) has significantly more tornadoes (41%) than elsewhere in the state and the MOB CWA (Mobile, AL) has significantly fewer tornadoes (53%). In South Dakota the FSD CWA (Sioux Fall, SD) has significantly more tornadoes (66%) than elsewhere in the state and the UNR CWA (Rapid City, SD) has significantly fewer tornadoes (34%). These difference, especially for South Dakota, likely reflect real differences in climatology rather than differences in warning and verification procedures.

5. Summary and Future Directions

Tornadoes are discrete events, clustered in space and time, and locally quite rare. This makes it difficult to construct a regional climatology. Here a statistical model is demonstrated that overcomes some of these difficulties and that produces a smoothed regional-scale climatology of tornado occurrences. The model is applied to data aggregated to the county level. Data consist of annual population and tornado counts as well as an index of terrain roughness derived from a digital elevation model. The statistical model includes a term that represents the smoothed frequency relative to the state average after accounting for changes in reporting from population shifts and from improvements in rating procedures. The model is Bayesian and is fit using the method of integrated nested Laplacian approximation (INLA). A map of the correlated random-effects term shows where tornado activity is high relative to the state average. The model is used to check whether high-resolution variation in terrain elevation is related to tornado frequency and whether there are differences in tornado activity by CWA.

The data preparation and model-fitting procedures were described using data from Kansas over the period 1970–2013. A key finding is that Kansas tornado reports increase by 13% with a two-fold increase in population but the influence of population density is decreasing. Independent of this relationship tornadoes have been increasing at an annual rate of 1.9%. Another key finding is the significant pattern of correlated residuals showing more Kansas tornadoes in a corridor of counties running roughly north to south across the west central part of the state. The model is improved by adding a term indexing terrain roughness. The magnitude of this effect, estimated by the posterior mean of the coefficient, amounts to an 18% reduction in the number of tornadoes for every ten meter increase in elevation standard deviation. The model indicates that tornadoes are 51% more likely to occur in counties served by the CWAs of DDC and GID as elsewhere in the state.

Flexibility of the model was illustrated by fitting it to data from other tornado-prone states including Illinois, Mississippi, South Dakota, and Ohio. Population changes are an important term especially in South Dakota and Mississippi. In Mississippi the model indicates a 20% increase in tornado reports for a doubling of the population. A significant downward trend at a rate of 1.7% per year is noted in the South Dakota tornado model and a significant upward trend at a rate of 2.4% per year is noted in the Mississippi tornado model. The Brier score is lowest for the Ohio model.

Terrain roughness is a significant explanatory factor for Mississippi tornadoes and a marginally significant factor for South Dakota tornadoes, but not for tornadoes elsewhere. Across Mississippi the magnitude of the roughness effect amounts to a 10% reduction in tornadoes for every ten meter increase in elevation standard deviation. The CWAs are not a significant factor in explaining the pattern of tornadoes in Illinois and Ohio. However in Mississippi the Jackson CWA sees 41% more tornadoes on average than elsewhere in the state. In South Dakota the Sioux Falls CWA sees 66% more tornadoes than elsewhere in the state. These spatial variations likely reflect real differences in tornado climatology rather than differences in warning and verification procedures.

Future studies will test additional hypotheses. For example, is the influence of roughness less for the subset of strongest tornadoes? The model will also be extended to include other local and regional variables including land use and land cover. Of particular interest is a test of the physical hypothesis that gradients in soil moisture contribute to tornado genesis (Lanicci et al. 1987). Interest also centers on using the adjusted tornado counts as the actual risk of tornadoes together with demographic and social data to examine regions most vulnerable to tornadoes. The model can be extended to include multiple states and it can be adapted for use with a regular grid. The model can also be adjusted for other tornado data. For example, it might be interesting to use tornado path length as the response variable rather than tornado count. Path length provides a better metric for the influence a tornado has on a region (Dixon et al. 2014).

Acknowledgments. Partial financial support for this research came from Climatek.

References

Anderson, C. J., C. K. Wikle, and Q. Zhou, 2007: Population influences on tornado reports in the United States. *Wea. Forecasting*, **22**, 571–579.

Besag, J., 1975: Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **24**, 179–195.

Bivand, R., 2014: *spdep: Spatial dependence: weighting schemes, statistics and models*. URL http://CRAN.R-project.org/package=spdep, r package version 0.5-74.

Blangiardo, M., M. Cameletti, G. Baio, and H. Rue, 2013: Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-temporal Epidemiology*, **4 (0)**, 33 – 49, doi:http://dx.doi.org/10.1016/j.sste.2012.12.001, URL http://www.sciencedirect.com/science/article/pii/S1877584512000846.

Brooks, H., and C. Doswell, 2001: Normalized damage from major tornadoes in the United States: 18901999. *Wea. Forecasting*, **16**, 168–176.

Brooks, H. E., 2013: Severe thunderstorms and climate change. *Atmos. Res.*, **123**, 129 – 138.

Brooks, H. E., C. A. Doswell, and M. P. Kay, 2003: Climatological estimates of local daily tornado probability for the United States. *Wea. Forecasting*, **18**, 626–640.

Church, C. R., J. T. Snow, G. L. Baker, and E. M. Agee, 1979: Characteristics of tornado like vortices as a function of swirl ratio: A laboratory investigation. *J. Atmos. Sci.*, **36**, 1755–1776.

Coleman, T. A., and P. G. Dixon, 2014: An objective analysis of tornado risk in the United States. *Wea. Forecasting*, 366–376.

Czado, C., T. Gneiting, and L. Held, 2009: Predictive model assessment for count data. *Biometrics*, **65**, 1254–1261.

Davies-Jones, R. P., 1973: The dependence of core radius on swirl ratio in tornado simulator. *J. Atmos. Sci.*, **30**, 1427–1430.

Dessens, J., 1972: Influence of ground roughness on tornadoes: A laboratory simulation. *J. Appl. Meteor.*, **11**, 72–75.

Diffenbaugh, N. S., R. J. Trapp, and H. Brooks, 2008: Does global warming influence tornado activity. *EOS, Transactions*, **89 (53)**, 553–560.

Dixon, P. G., A. E. Mercer, J. Choi, and J. S. Allen, 2011: Tornado risk analysis: Is Dixie alley an extension of tornado alley? *Bull. Amer. Meteor. Soc.*, **92**, 433–441.

Dixon, P. G., A. E. Mercer, K. Grala, and W. H. Cooke, 2014: Objective identification of tornado seasons and ideal spatial smoothing radii. *Earth Interactions*, **18**, 1–15.

Doswell, C., 2007: Small sample size and data quality issues illustrated using tornado occurrence data. *Electronic Journal of Severe Storms Meteorology*, **116 (2)**, 1–10.

Doswell, C. A., H. E. Brooks, and M. P. Kay, 2005: Climatological estimates of daily local non-tornadic severe thunderstorm probability for the United States. *Wea. Forecasting*, **20**, 577–595.

Doswell, C. A., and D. W. Burgess, 1988: On some issues of United States tornado climatology. *Mon. Wea. Rev.*, **116**, 495–501.

Doswell, C. A., A. R. Moller, and H. E. Brooks, 1999: Storm spotting and public awareness since the first tornado forecasts of 1948. *Wea. Forecasting*, **14**, 544–557.

Edwards, R., J. G. LaDue, J. T. Ferree, K. Scharfenberg, C. Maier, and W. L. Coulbourne, 2013: Tornado intensity estimation: Past, present, and future. *Bull. Amer. Meteor. Soc.*, **94**, 641–653.

Elsner, J. B., and T. H. Jagger, 2013: *Hurricane Climatology: A Modern Statistical Guide Using R*. Oxford University Press, USA.

Elsner, J. B., L. E. Michaels, K. N. Scheitlin, and I. J. Elsner, 2013: The decreasing population bias in tornado reports. *Weather, Climate, and Society*, **5**, 221–232.

Elsner, J. B., and H. M. Widen, 2014: Predicting spring tornado activity in the central Great Plains by March 1st. *Mon. Wea. Rev.*, **142**, 259–267.

Fujita, T., and A. D. Pearson, 1973: Results of FPP classification of 1971 and 1972 tornadoes. *Eight Conference on Severe Local Storms*, 142-145.

Gneiting, T., and A. Raftery, 2007: Strictly proper scoring rules, prediction and estimation. *Journal of the American Statistical Association, Series B*, **102**, 359–378.

Karpman, D., M. A. R. Ferreira, and C. K. Wikle, 2013: A point process model for tornado report climatology. *Stat*, **2**, 1–8.

King, P., 1997: On the absence of population bias in the tornado climatology of southwestern Ontario. *Wea. Forecasting*, **12**, 939–946.

Lanicci, J. M., T. Carlson, and T. Warner, 1987: Sensitivity of the great plains severe-storm environment to ski-moisture distribution. *Mon. Wea. Rev.*, **115**, 2660–2673.

Leslie, F. W., 1977: Surface roughness effects on suction vortex formation: A laboratory simulation. *J. Atmos. Sci.*, **34**, 1022–1027.

Lewellen, W. S., 1962: A solution for 3 dimensional vortex flows with strong circulation. *Journal of Fluid Mechanics*, **14**, 420–432.

Potter, S., 2007: Fine-tuning Fujita. *Weatherwise*, **60**, 64–71.

Ramsdell, J. V., Jr, and J. P. Rishel, 2007: Tornado Climatology of the Contiguous United States. Tech. Rep. NUREG/CR-4461, PNNL-15112, Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352.

Ray, P. S., P. Bieringer, X. Niu, and B. Whissel, 2003: An improved estimate of tornado occurrence in the central Plains of the United States. *Mon. Wea. Rev.*, **131**, 1026–1031.

Rue, H., S. Martino, and N. Chopin, 2009: Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **71**, 319–392.

Rue, H., S. Martino, F. Lindgren, D. Simpson, A. Riebler, and E. T. Krainski, 2014: *INLA: Functions which allow to perform full Bayesian analysis of latent Gaussian models using Integrated Nested Laplace Approximaxion*. R package version 0.0-1401811493.

Schaefer, J. T., and R. Edwards, 1999: The SPC tornado/severe thunderstorm database. *11th Conference on Applied Climatology*.

Schrödle, B., and L. Held, 2011: Spatio-temporal disease mapping using inla. *Environmetrics*, **22 (6)**, 725–734, doi:10.1002/env.1065, URL http://dx.doi.org/10.1002/env.1065.

Schultz, D. M., Y. P. Richardson, P. M. Markowski, and C. A. Doswell, 2014: Tornadoes in the central United States and the "Clash of Air Masses". *Bull. Amer. Meteor. Soc.*, **95**, xx–xx, doi: http://dx.doi.org/10.1175/BAMS-D-13-00252.1.

Schultz, D. M., C. C. Weiss, and P. M. Hoffman, 2007: The synoptic regulation of dryline intensity. *Monthly Weather Review*, **135**, 1699–1709.

Shafer, C. M., and C. A. Doswell, 2011: Using kernel density estimation to identify, rank, and classify severe weather outbreak events. *Electronic Journal of Severe Storms Meteorology*, **6**, 1–28.

Snider, C. R., 1977: A look at Michigan tornado statistics. *Mon. Wea. Rev.*, **105**, 1341–1342.

Verbout, S. M., H. E. Brooks, L. M. Leslie, and D. M. Schultz, 2006: Evolution of the U.S. tornado database: 1954-2003. *Wea. Forecasting*, **21**, 86–93.

Wilson, J. W., and S. A. Changnon, 1971: *Illinois tornadoes*. Illinois Department of Registration and Education, Illinois State Water Survey.

LIST OF TABLES

TABLE 1. Summary of the data analysis and modeling results. DIC is the deviance information criterion, AD is the Anderson-Darling test, and r is the Pearson correlation coefficient.

	Kansas	Illinois	Mississippi	South Dakota	Ohio
FIPS	20	17	28	46	39
No. counties	105	102	82	66	88
Area (km^2)	210,845	144,451	123,701	199,367	105,954
Avg Elevation [m]	580.9 (580.5,581.2)	189.1 (189.0,189.2)	85.70 (85.62,85.78)	665.0 (664.6,665.4)	279.6 (279.5,279.7)
No. tornadoes (nT) [EF1+]	1010	879	1112	423	501
r(Area, nT)	.34 (.19,.48)	.64 (.53,.72)	.55 (.41,.67)	−.11 (−.30,.10)	.34 (.17,.49)
Single tornado most counties	7	8	12	3	6
Population [2012]	2,893,957	12,882,135	2,991,207	833,354	11,544,225
r(Population [2012], nT)	.04 (−.12,.20)	.14 (−.02,.30)	.49 (.34,.62)	.39 (.20,.55)	.20 (.03,.37)
Tornado trend [%/yr]	.87 (−.27,2.0)	.48 (−.77,1.75)	.44 (−.78,1.67)	−1.60 (−3.04,−.14)	−1.45 (−2.85,−.03)
DIC (w/out spatial term)	6027	5268	5729	2544	3364
DIC (w/ spatial term)	5990	5211	5680	2500	3302
AD p value	>.15	>.15	>.15	>.15	>.15
Log score	.635	.568	.770	.448	.436
Brier score	.570	.415	.564	.269	.212
Pop density term	.1187 (.0655,.1723)	.1083 (.0525,.1643)	.1304 (.0734,.1868)	.1693 (.0791,.2569)	.0714 (−.0051,.1466)
Trend term	.0189 (.0054,.0323)	n.s.	.0230 (.0039,.0422)	−.0173 (−.0318,−.0029)	n.s.
Interaction term	−.0045 (−.0073,−.0017)	−.0016 (−.0036,.0004)	−.0050 (−.0083,−.0018)	n.s.	−.0031 (−.0053,−.0010)
r(Roughness, nT)	−.067 (−.256,.126)	−.173 (−.355,.022)	−.023 (−.239,.195)	−.115 (−.347,.131)	−.066 (−.271,.146)
DIC (w/ Roughness term)	5980	5212	5678	2502	3303
Roughness term	−.0186 (−.0268,−.0106)	−.0051 (−.0173,.0073)	−.0098 (−.0194,−.0003)	−.0020 (−.0039,.0000)	.0003 (−.0126,.0133)
DIC (w/ CWA term)	5981	5213	5669	2881	3352

LIST OF FIGURES

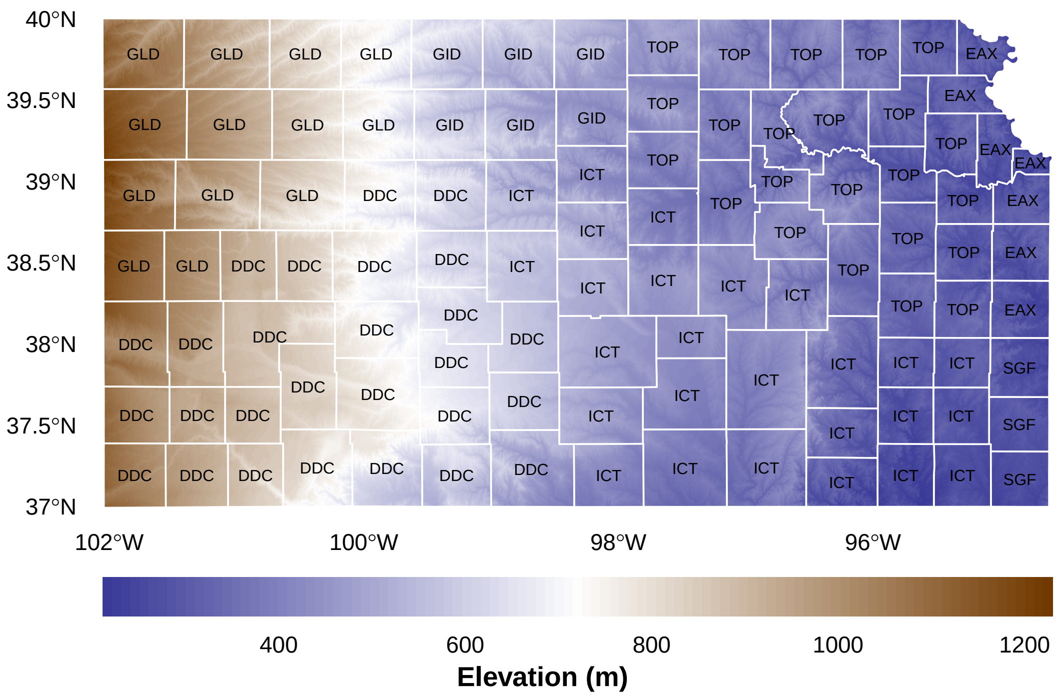

FIG. 1. Kansas counties and elevation. Counties are labeled by the corresponding CWA. Elevation is given at a resolution of 80 m.

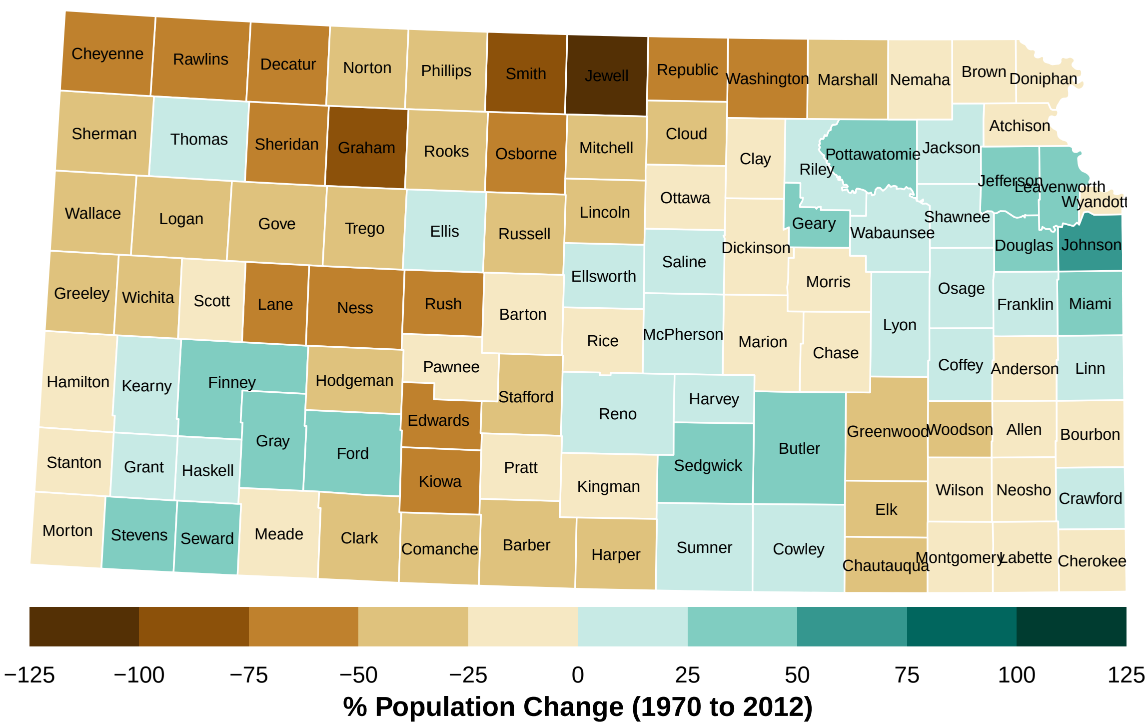

FIG. 2. Population changes between 1970 and 2012. The change is expressed as a percentage difference with 2012 as the base year.

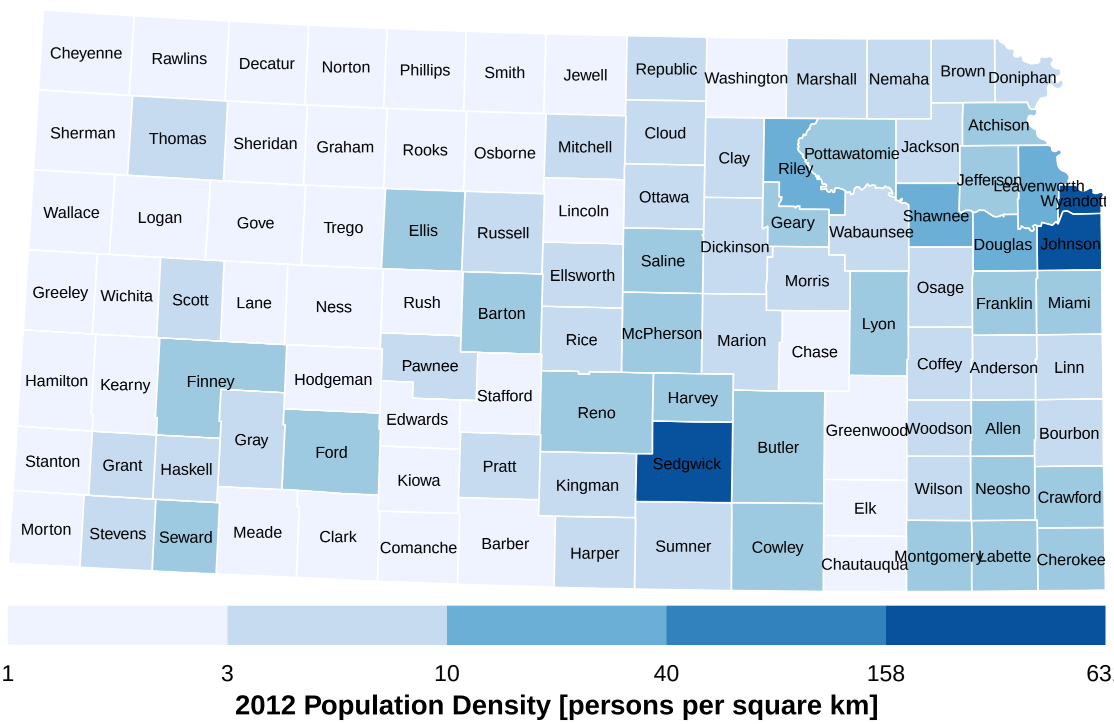

FIG. 3. Population estimates for 2012 by county. Values are expressed as persons per square km.

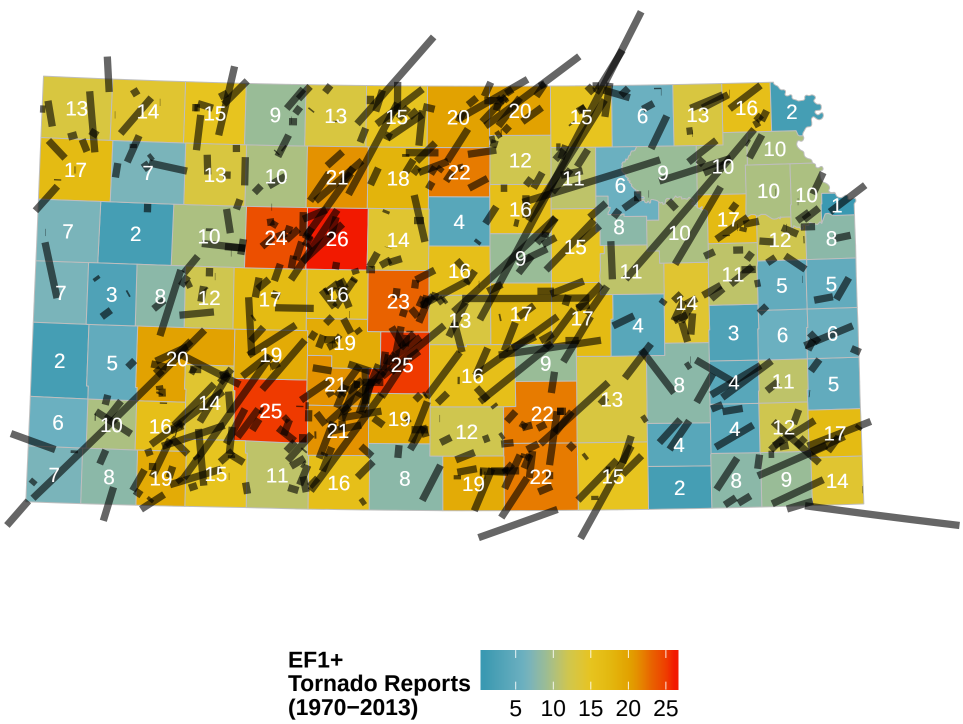

FIG. 4. Tornado report frequency by county for Kansas. Only tornadoes rated EF1 and higher are used. Lines show the tornado track. The shortest tracks are not visible at this scale. Total tornado counts over the period 1970–2013 are listed inside the county and the color scale is from few (blue) to many (red).

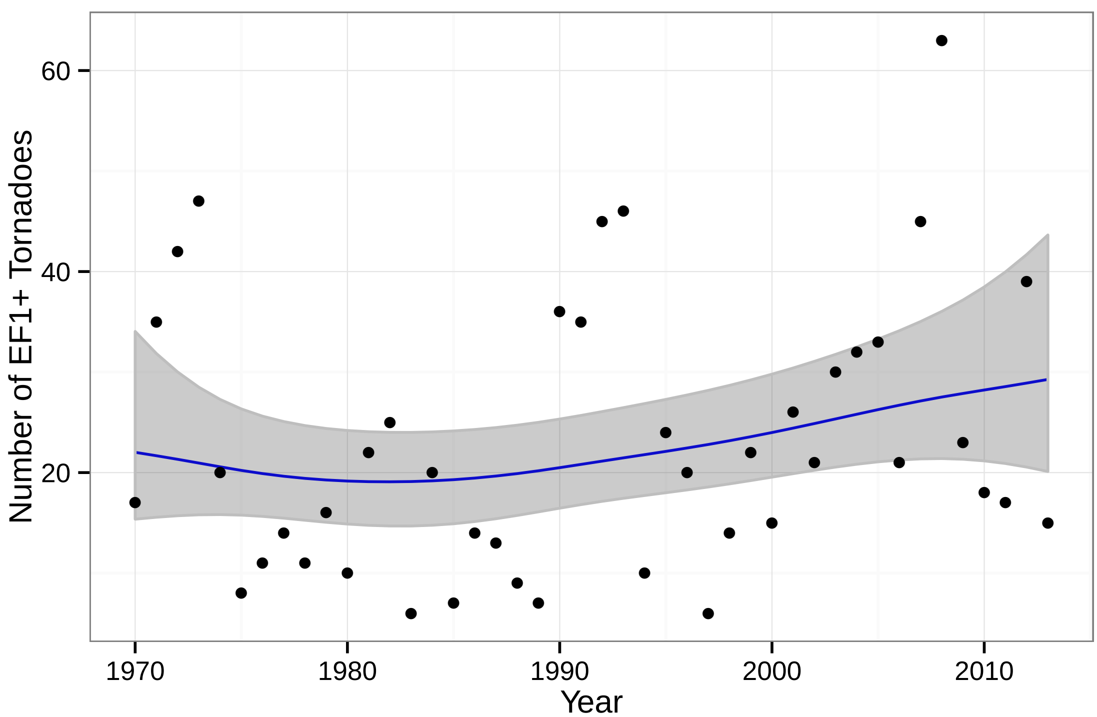

FIG. 5. Statewide tornado counts for Kansas from 1970–2013. The trend line uses a second-order random walk model where the counts are described by a negative binomial distribution. The 90% uncertainty band is shown in gray.

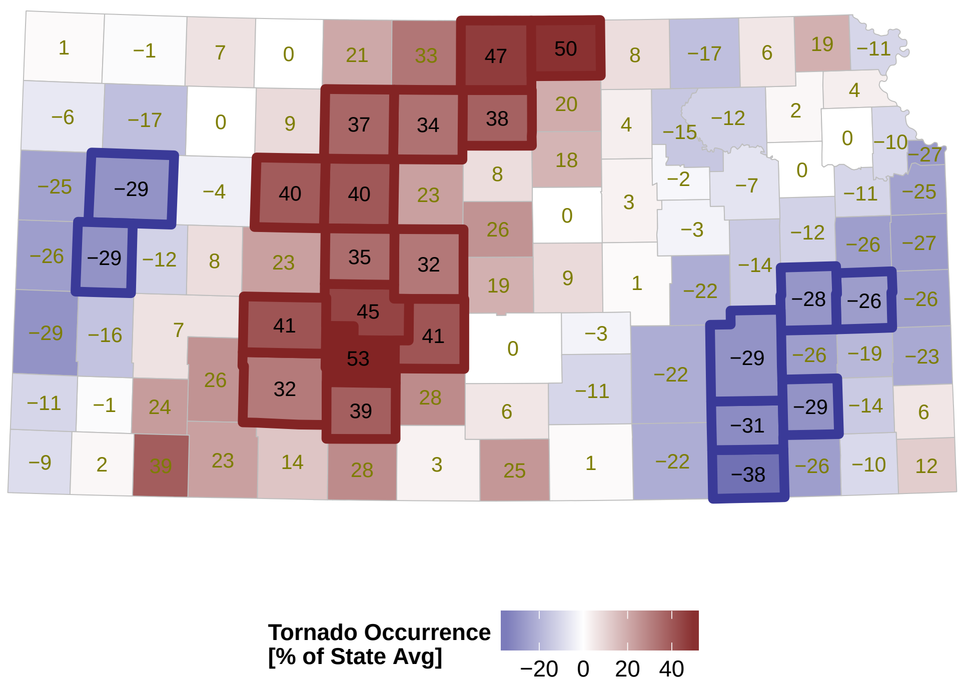

FIG. 6. Correlated random effects from the Kansas tornado model. Values are the posterior mean and are expressed as the percent difference from the state average. The model includes annual population density and calendar year as fixed effects.

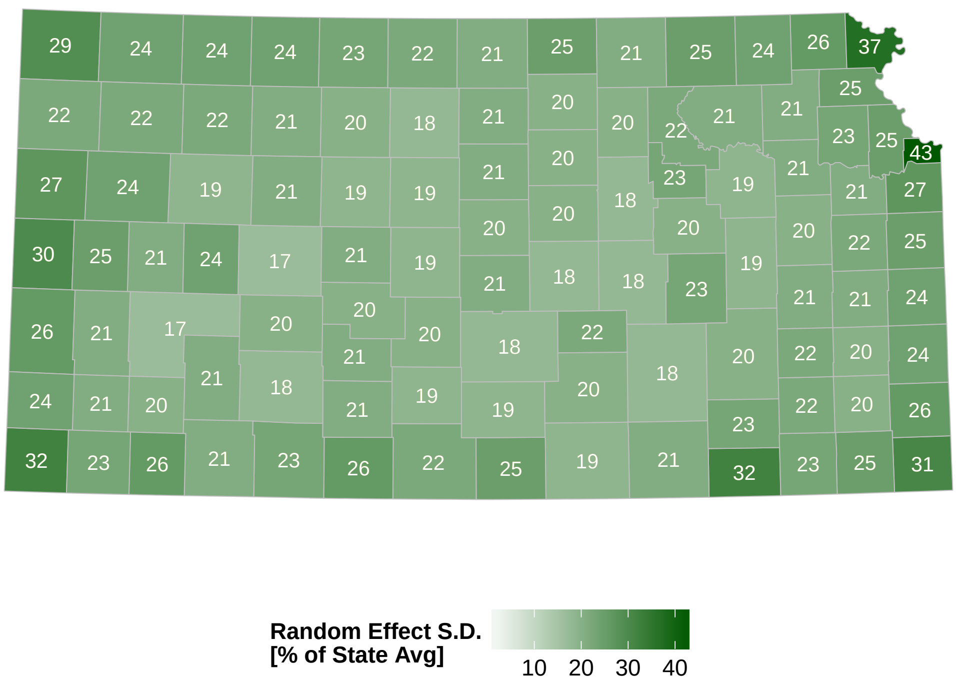

FIG. 7. Standard deviation of the correlated random effects from the Kansas tornado model. Values have units of percent difference from the state average.

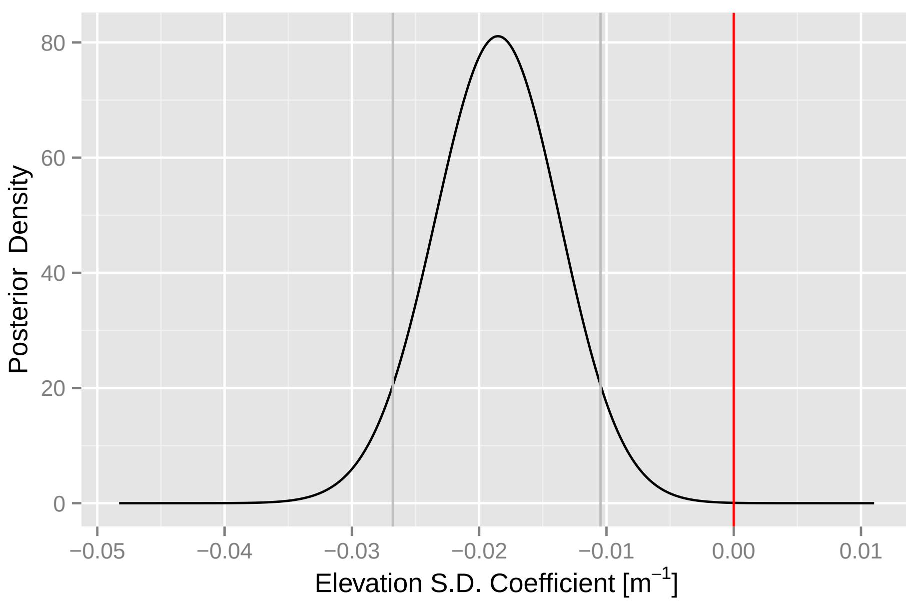

FIG. 8. Posterior density of the elevation standard deviation term. The 90% CI is shown with the vertical gray lines. The red line indicates no effect.

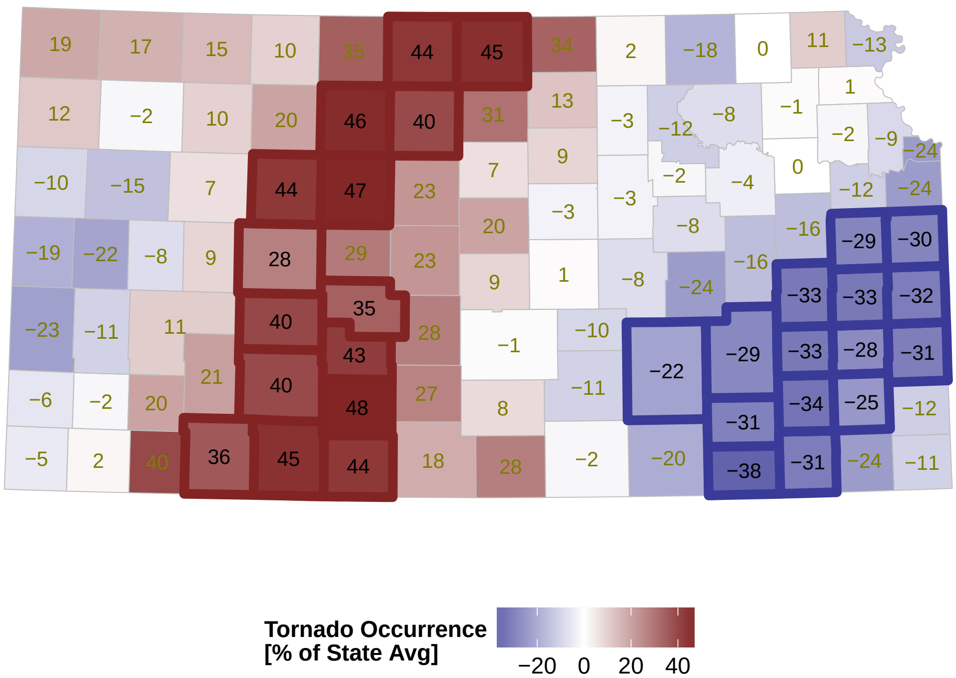

FIG. 9. Same as Fig. 6 except the model has elevation standard deviation as an additional fixed effect.

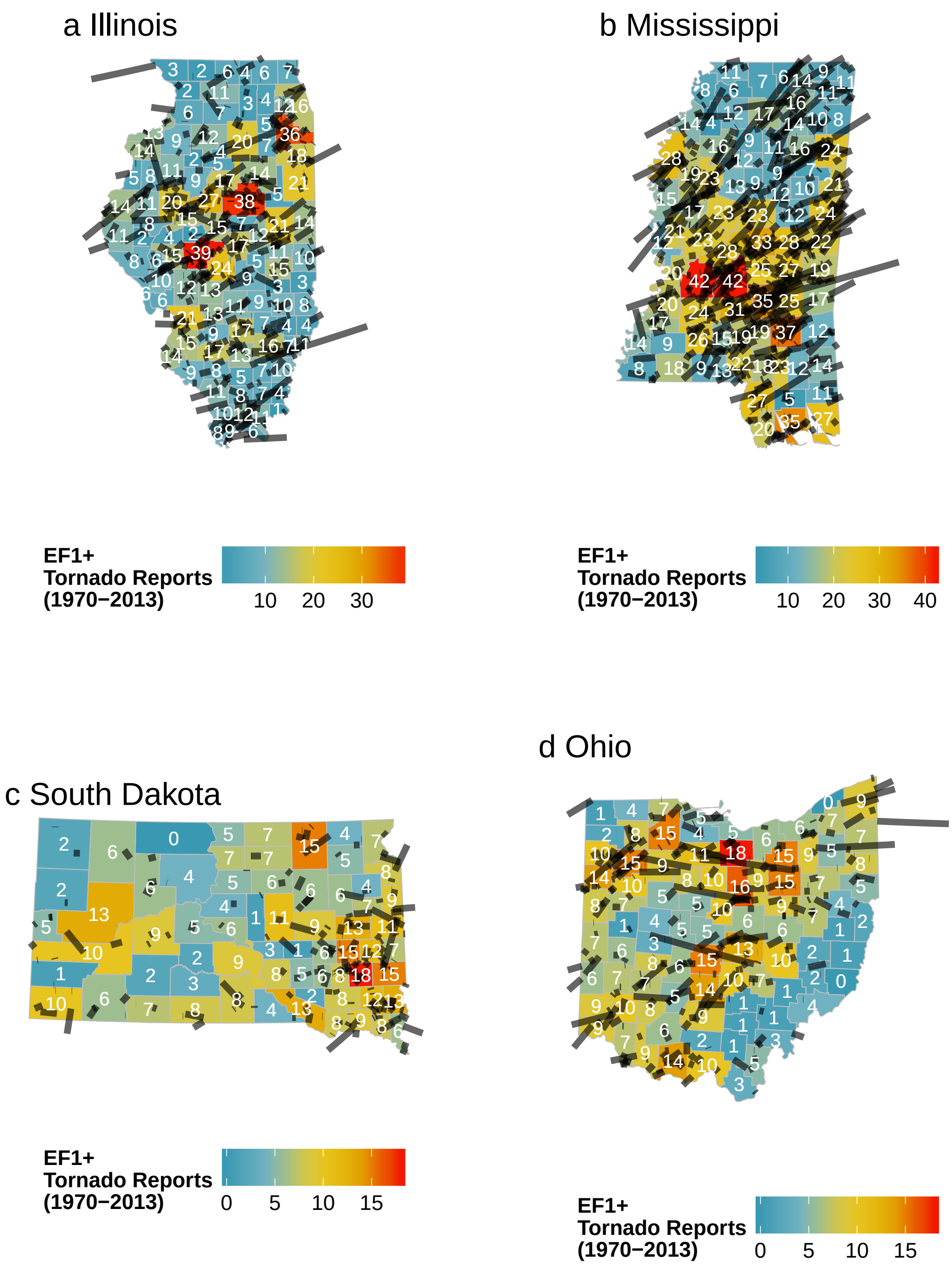

FIG. 10. Tornado report frequency by county for Illinois, Mississippi, South Dakota, and Ohio.

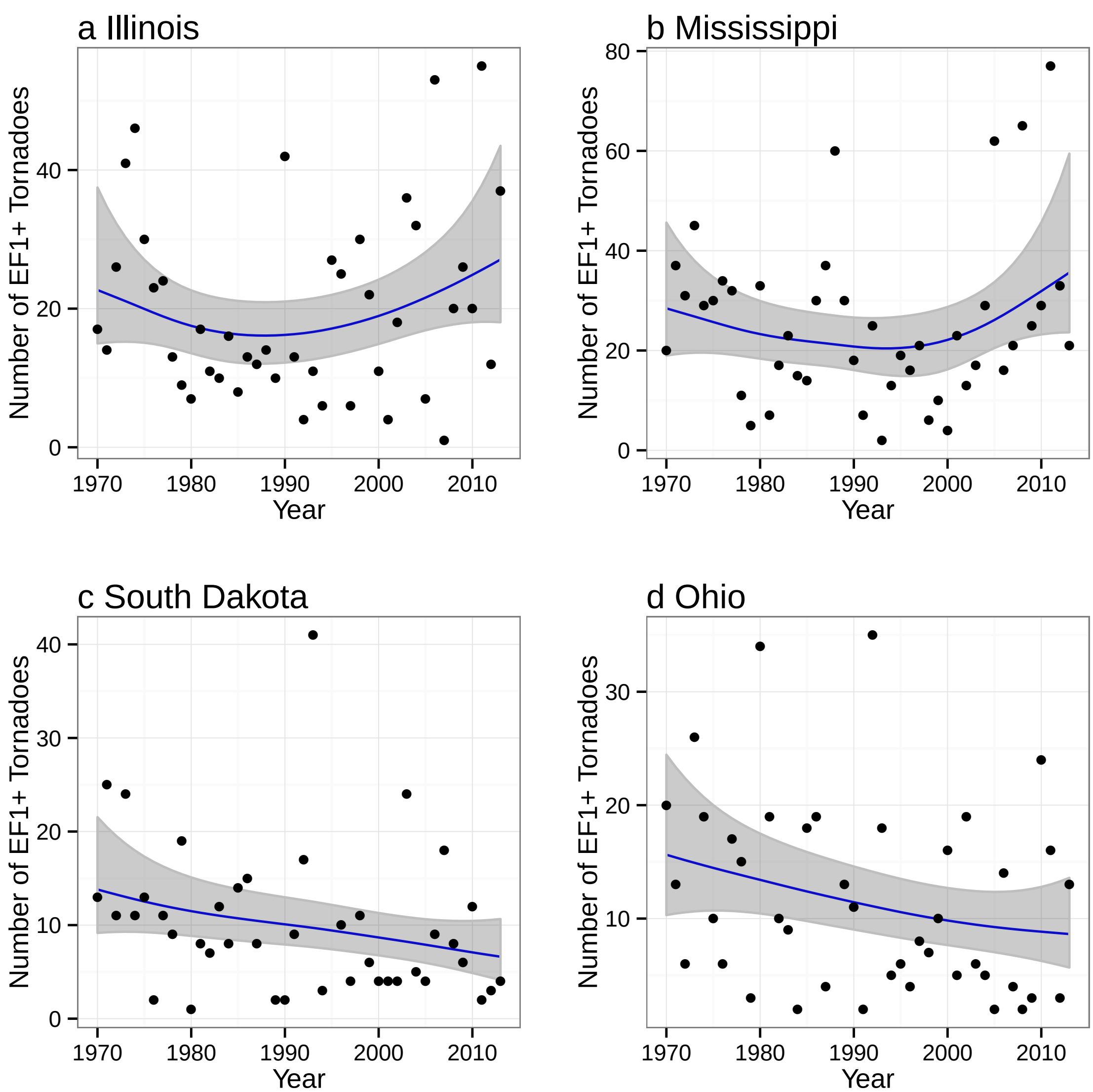

FIG. 11. Statewide tornado counts.

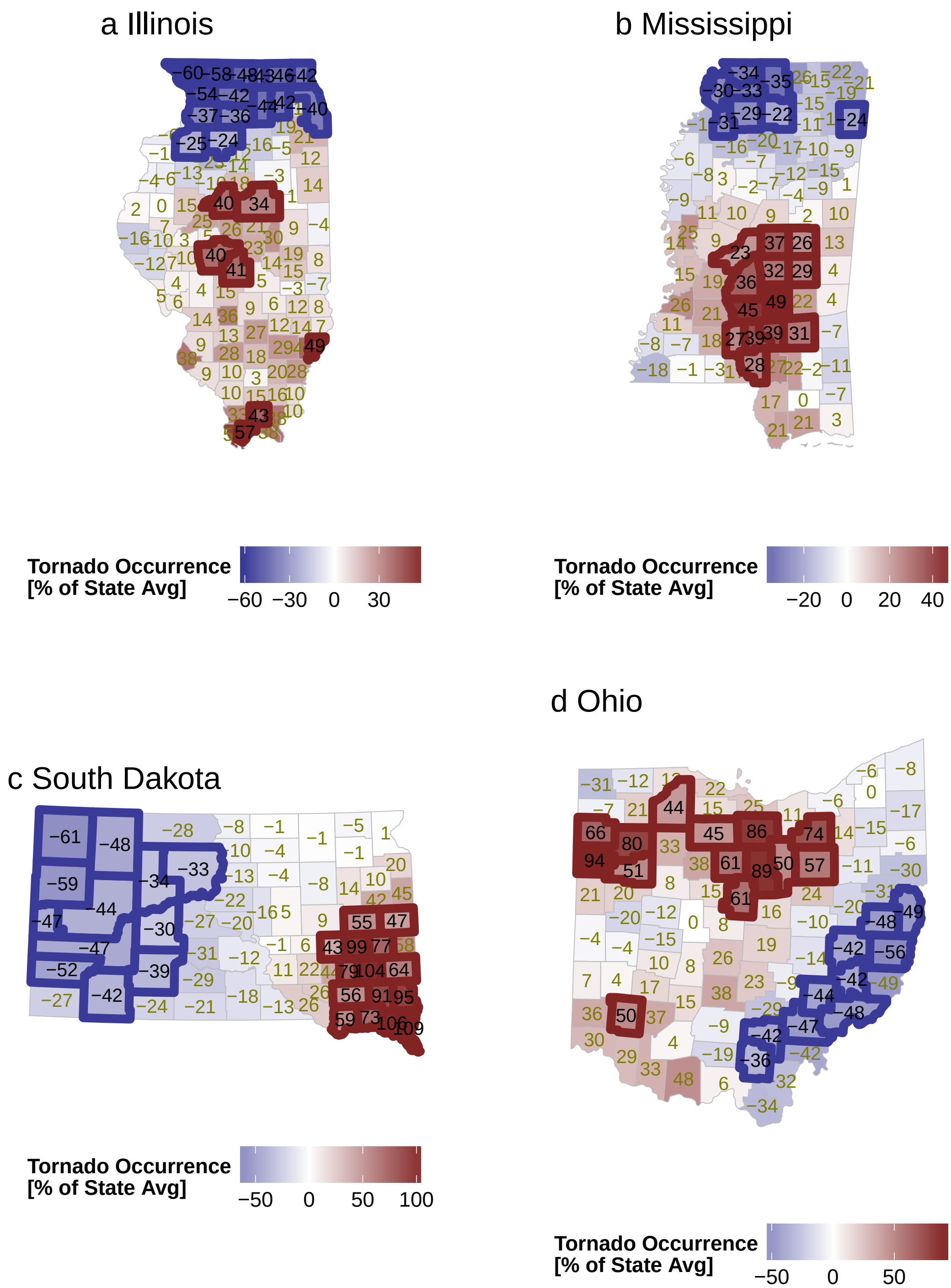

FIG. 12. Correlated random effects from the state tornado models.

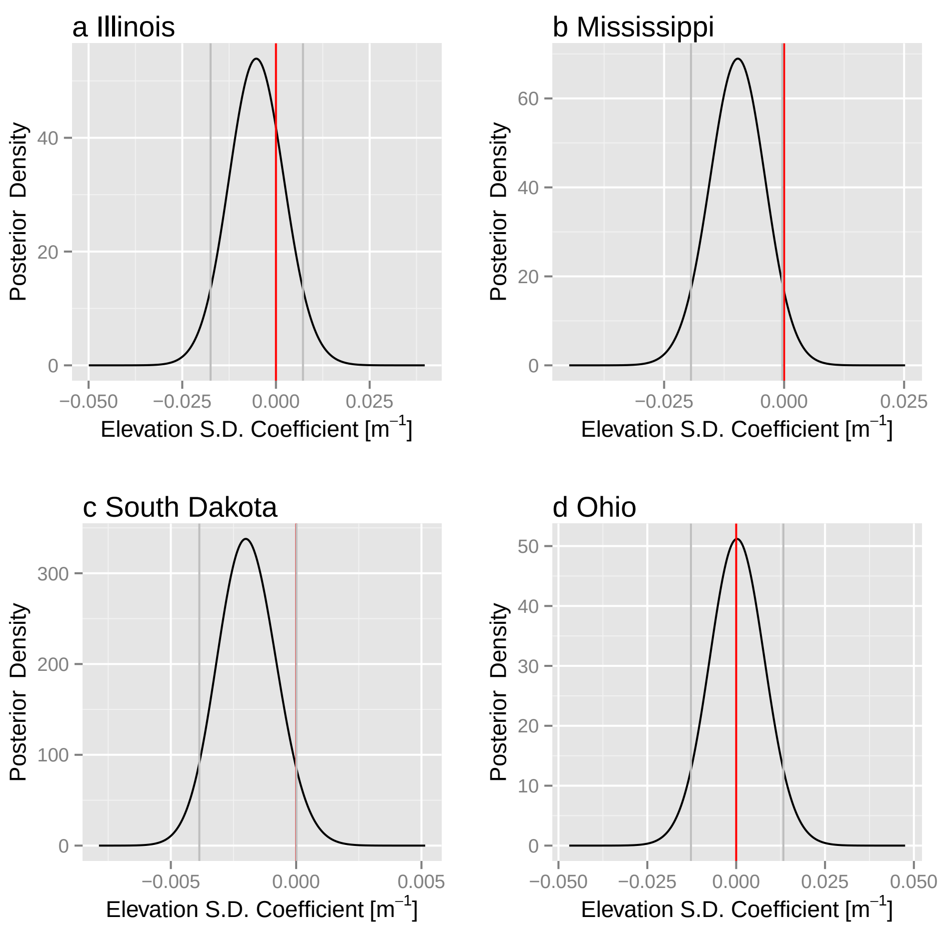

FIG. 13. Posterior density of the elevation standard deviation term from the state tornado models.